Maternal genome-wide DNA methylation profiling in gestational diabetes shows distinctive disease-associated changes relative to matched healthy pregnancies

Dr Pensee Wu[1,2], Prof William E. Farrell[1], Dr Kim E. Haworth[1], Prof Richard D. Emes[3,4], Dr Mark O. Kitchen[1], Dr John R. Glossop[1,5], Dr Fahmy W. Hanna[6], Prof Anthony A. Fryer[1]

[1]Institute for Science and Technology in Medicine; Keele University; Guy Hilton Research Centre; Staffordshire, UK

[2]Academic Unit of Obstetrics and Gynaecology, University Hospital of North Midlands NHS Trust, Stoke-on-Trent, Staffordshire, UK

[3]School of Veterinary Medicine and Science; University of Nottingham; Leicestershire, UK

[4]Advanced Data Analysis Centre; University of Nottingham; Leicestershire, UK

[5]Haywood Rheumatology Centre; Haywood Hospital; Staffordshire, UK

[6]Department of Diabetes and Endocrinology, University Hospital of North Midlands NHS Trust, Stoke-on-Trent, Staffordshire, UK

All listed authors have no conflicts-of-interest to declare, and have no relevant financial relationships to disclose.

Abbreviations

DNA	Deoxyribonucleic acid
EDTA	Ethylenediaminetetraacetic acid
GDM	Gestational diabetes
KEGG	Kyoto encyclopedia of genes and genomes
T2DM	Type 2 diabetes mellitus
QUIN	Quinolinic acid
SNP	Single-nucleotide polymorphism
SWAN	Subset-quantile within array normalisation

Abstract

Several recent reports have described associations between gestational diabetes (GDM) and changes to the epigenomic landscape where the DNA samples were derived from either cord or placental sources. We employed genome-wide 450k BeadChip-Array analysis to determine changes to the epigenome in a unique cohort of maternal blood DNA from 11 pregnant women prior to GDM development relative to matched controls. Hierarchical clustering segregated the samples into two distinct clusters comprising GDM and healthy pregnancies. Screening identified 100 CpGs with a mean β-value difference of ≥0.2 between cases and controls. Using stringent criteria, 5 CpGs (within *COPS8*, *PIK3R5*, *HAAO*, *CCDC124* and *C5orf34* genes), demonstrated potentials to be clinical biomarkers as revealed by differential methylation in 8 of 11 women who developed GDM relative to matched controls. We identified, for the first time, maternal methylation changes *prior* to the onset of GDM that may prove useful as biomarkers for early therapeutic intervention.

Key words: gestational diabetes, epigenetics, fetal programming, biomarker, 450k array

Introduction

Gestational diabetes (GDM) is a pregnancy-specific endocrine disorder with a prevalence of 3.5-14%.[1] Due to the worldwide obesity epidemic and recently modified diagnostic criteria, GDM is increasingly prevalent.[2] It occurs because of a mismatch between insulin production and requirement, leading to maternal hyperglycaemia. Since glucose is able to cross the placenta, whilst insulin does not, the fetus is also exposed to hyperglycaemic conditions. Women with GDM are at increased risk of Caesarean section and stillbirth compared with healthy women.[3, 4] They are also more likely to develop type 2 diabetes (T2DM), dyslipidemia and cardiovascular disease in later life,[5-7] while their offspring have an increased long-term risk of obesity and diabetes.[2]

Epigenetic modifications, which may be causal of or associated with changes in gene expression, offer significant promise for understanding the underlying mechanisms of GDM. Indeed, and as an example, epigenetic changes in T2DM have been reported in genes involved in metabolism.[8-13] Since maternal epigenetic modifications are known to contribute to fetal programming,[14] recent studies have investigated the role of epigenetic alterations in offspring exposed to maternal hyperglycaemia and found positive associations.[15-19] Furthermore, previous studies suggest that epigenetic

67 modifications may play a role in the pathogenesis of GDM.[20, 21]

68

69 Epigenetic research in GDM has largely used targeted (candidate gene)
70 approaches.[15, 16, 18, 19] To date, only two studies have utilised genome-wide
71 methodology[17, 22] and in these cases investigators examined placenta and cord blood
72 samples from GDM pregnancies. Differentially methylated genes were identified
73 between GDM and healthy pregnancies,[17, 22] which provide evidence for the
74 involvement of these genes and/or their differential methylation in GDM. However,
75 there have been no genome-wide studies examining methylation differences between
76 *maternal* tissue samples from GDM and healthy pregnancies. We decided to focus on
77 maternal epigenetic profiles, as they would facilitate the assessment of the *in utero*
78 environment and allow identification of predictive biomarkers that would enable
79 targeted intervention to high risk groups.

80

81 On the basis of the current literature, we hypothesised the presence of pre-existing
82 epigenetic markers in women who subsequently go on to develop GDM. In this study,
83 and for the first time in this disease, we interrogated genome-wide DNA methylation in
84 peripheral blood samples collected from women *prior to* the development of GDM and
85 relative to matched healthy controls that did not develop GDM. Using this discovery

cohort, our aim was to identify candidate genes with future promise as potential biomarkers for the prediction of GDM in early pregnancy.

Results

Our initial data analyses focused on comparison of our data in antenatal samples with the two recent genome-wide studies that investigated cord blood and placental tissue samples.[17, 22] We compared our data with those of Finer *et al*.[22] and Ruchat *et al*.[17] separately due to the different approaches used for data processing by each study (**Figure 1**). Using the filtering criteria shown in step 1A of **Figure 1**, comparison of our data with those of Finer *et al*.[22] identified 4,755 differentially methylated CpGs (representing 2,236 genes) where the mean β-value difference between the GDM and healthy groups was >0.05 and statistically significant (p<0.05). In contrast, comparison with the data of Ruchat *et al*.[17] (step 1B of **Figure 1**) identified 1,035 CpGs (representing 633 candidate genes). We also performed the same comparison after applying multiple testing adjustment using the false discovery rates, which showed no overlap of our data with these two studies.

As shown in **Figure 2A**, by comparing the 2,236 genes identified as differentially methylated in our study with those reported by Finer *et al.*,[22] two genes were common

between maternal blood, umbilical cord and placenta: Hook Microtubule-Tethering Protein 2 (*HOOK2*) and Retinol Dehydrogenase 12 (*RDH12*). Conversely, and as summarised by the Venn diagram in **Figure 2B**, there were no genes common to all three tissue types when we compared our data with that of Ruchat *et al*.[17]

The 4,755 CpGs initially identified as differentially methylated were then subjected to further filtering (steps 2 and 3, **Figure 1**). Using this approach, we identified 100 unique CpGs (comprising 66 genes) that were differentially methylated between GDM and healthy pregnancies (the full annotated list is shown in **Table S1**). None of these CpGs have an annotated single-nucleotide polymorphism (SNP) in the probe. Closer examination of the 100 CpGs revealed that the majority (53%) were hypomethylated in GDM relative to healthy pregnancies. The observed differences in mean β-value showed a maximum difference of 0.38. The frequency and DNA methylation of these differentially methylated CpG sites in relation to their genomic location and CpG islands are shown in **Figure S1**. Of the differentially methylated CpGs, 45% were associated with a CpG island, shelf or shore (**Figure S1C)**.

Hierarchical clustering was performed to determine whether the methylation patterns in these 100 CpGs can be used to distinguish between GDM and healthy pregnancies.

The heatmap in **Figure 3** illustrates that there are distinctive methylation patterns between GDM and healthy pregnancies, which segregate samples into two distinct groups comprising those from GDM and healthy populations. The slide type did not cause the clustering, therefore our results were not due to batch effects. Calculation of the genomic inflation factor before and after normalisation steps showed that removal of SNP containing probes and subset-quantile within array normalisation (SWAN) normalisation by the *minfi* package reduced the genomic inflation.[23-25] Pre-normalisation λ was estimated to be 1.189 (standard error of the estimation = 9.461×10^{-5}) and after normalisation the estimated lambda was reduced to 1.132 (standard error of the estimation = 7.461×10^{-5}). The remaining genomic inflation suggests that mild confounding stratification factors remain unaccounted for in the data.

Enrichment of gene ontology terms and biological pathways within the 66 genes associated with differentially methylated CpGs were assessed using DAVID online software[26] identified 11 overrepresented pathways, with the top three (ranked by *p* value) involved in cell adhesion molecules, type 1 diabetes mellitus and keratin pathways. However, enrichment of these pathways was not statistically significant following adjustment for false discovery rates (**Table S2**).

Finally, we examined the absolute β-value differences across all 11 matched pairs. Using this stringent criteria, in 5 of the 100 CpGs identified, at least 8 of the 11 GDM pregnancies showed β-value differences of >0.2 relative to matched controls. The 5 CpGs comprised of 5 genes (*COPS8*, *PIK3R5*, *HAAO*, *C5orf34* and *CCDC124*) and their functions are shown in **Table 1**.

Discussion

We describe for the first time, genome-wide DNA methylation changes in maternal blood *prior* to the diagnosis of GDM. We identified 2 differentially methylated genes that shared identity with genes previously described in studies which interrogated placenta and umbilical cord blood samples and, in these cases, using the same array platforms.[17, 22] Furthermore, using stringent filtering criteria, we identified 100 unique CpGs which segregated GDM and healthy pregnancies into distinct groups upon hierarchical clustering.

The strength of our study, in contrast to the previous studies, is that we carefully matched each GDM pregnancy to a healthy one to ensure the samples were comparable.[17, 22] Furthermore, as all samples were taken prior to development of

pregnancy complications, there was limited sampling bias.

We were able to compare our data to those from two recent genome-wide studies in GDM using cord blood and placenta tissue.[17, 22] Comparative analysis with Finer *et al*.[22] showed that *HOOK2* and *RDH12* were common to maternal blood, placenta and cord blood. *HOOK2* codes for a linker protein which mediates binding to organelles and is responsible for morphogenesis of cilia and endocytosis.[27, 28,29] *RDH12* encodes a retinal reductase, which also plays a role in the metabolism of short-chain aldehydes.[27, 30] In terms of KEGG orthology, it is involved in metabolic pathways as well as retinal metabolism.[31] These two genes, therefore, may represent important candidates for further study.

The disparity of candidate genes when comparisons are made to the previous studies might reflect the different data filtering criteria used by Ruchat *et al.* [17] and Finer *et al.*[22] Using the Finer *et al.* criteria, many of the differentially methylated CpGs are likely to have β-value differences of <0.2, which could be difficult to reproduce either by alternative methodologies such as Pyrosequencing or in replication studies using independent patient cohorts. Moreover, we used a distinct patient population to the other two studies. We used samples from women *prior* to the development of their

GDM, while both Ruchat *et al.*[17] and Finer *et al.*[22] used samples from women with established GDM. Furthermore, we used maternal blood samples, rather than placenta and cord blood samples. These disparities may have contributed to the differences in the absolute numbers of CpGs/genes identified.

Further analysis of our cohort identified 100 independent CpGs (comprising 66 genes), which were found to cluster GDM and healthy pregnancies separately. Reassuringly, these CpGs have no annotated SNPs in the probe. Enrichment of gene ontology terms and biological pathways of these 66 genes showed enrichment for genes involved in cell adhesion, type 1 diabetes mellitus and keratin pathways.[26, 32] Although the enrichment was not statistically significant following adjustment for false discovery rates, these are promising candidates which are worth examining to elucidate the biological mechanisms behind GDM. In future work, it will be important to verify, in larger independent cohorts, the candidates identified herein and to determine the impact of differential methylation. This may in the future improve the understanding of GDM pathogenesis and aid in the development of therapy.

The design of this pilot study was to generate a list of genes of interest using a relatively small number of samples. In order to avoid type II errors (false negatives),

200 we used uncorrected *p* values to identify potential candidates in the preliminary
201 screening. We then applied more stringent methodology (steps 2-4 of Figure 1) to
202 identify candidate genes. A potential limitation of our study is the possibility of
203 genomic inflation. Mild confounding stratification factors, such as changes in
204 composition of blood during the pregnancy, the time of blood sampling, and parity,
205 may have inflated the data. Therefore, we further validated the array data using an
206 independent method with Pyrosequencing in order to confirm our findings.
207
208 On closer inspection, 8 of 11 women who subsequently developed GDM showed
209 differential methylation at 5 CpGs (consisting of *COPS8*, *PIK3R5*, *HAAO*, *CCDC124*
210 and *C5orf34* genes) relative to matched controls. *COPS8* encodes a regulator of
211 multiple signaling pathways.[27, 33] It is involved in protein binding and negative
212 regulation of cell proliferation.[33, 34] The PIK3R5 protein has important roles in cell
213 growth, proliferation, motility, differentiation, survival, and intracellular trafficking.[27,
214 35-37] The HAAO protein catalyses the synthesis of quinolinic acid (QUIN). Increased
215 cerebral levels of QUIN may participate in the pathogenesis of neurologic and
216 inflammatory disorders, which may be mediated by HAAO.[27, 38] This unique epigenetic
217 signature may form the basis of future biomarker studies using a larger validation
218 cohort. The CCDC124 protein is involved in cell cycle and division.[39] *C5orf34* encodes

for a protein which is highly conserved across species, however its function remains uncharacterised.[27]

In summary, for the first time using a genome-wide approach in maternal blood, we have identified maternal methylation changes *prior* to the diagnosis of GDM. As a discovery-based study, our findings may prove useful towards developing simple biomarkers for predicting GDM, thus facilitating intervention strategies in the early antenatal period to improve the health of the mother and baby, both during pregnancy and in the long-term.

Materials and Methods

Patients

Peripheral blood samples were obtained from women prospectively recruited at the University Hospital of North Midlands, UK, between 12-16 weeks gestation, prior to the diagnosis of any pregnancy complications as part of the EFFECT-M study.[40] At the end of pregnancy, we identified 11 women who had GDM and individually matched each one with a healthy woman who had a normal pregnancy. They were matched in terms of age, body mass index, ethnicity, smoking status, medications and folate supplementation (**Table S3**). The study was approved by the West Midlands (Black

238 Country) Research Ethics Committee (REC reference no. 08/H1204/121).

239

240 Genome-wide DNA methylation profiling

241 We performed genome-wide analysis of DNA methylation using the Illumina
242 HumanMethylation450 BeadChip-Array which examines over 480,000 individual CpG
243 sites. We first extracted genomic DNA from blood samples collected into potassium
244 EDTA using standard phenol/chloroform procedures. Next, samples were sodium
245 bisulfite converted [41] and hybridised to arrays according to Illumina recommended
246 protocols that we have previously described.[42] Methylation at individual CpGs is
247 reported as a methylation β-value, which is a quantitative measure of methylation for
248 each CpG site with range between 0 (no methylation) to 1 (completely methylated).

249

250 Validation by sodium bisulfite pyrosequencing

251 A technical validation between array β-values and methylation levels was determined
252 by sodium bisulfite pyrosequencing in all 22 samples. To increase template quantity
253 for Pyrosequencing assays, whole genome amplification of bisulfite-converted DNA
254 followed by touchdown PCR were performed as previously described.[42] A PyroMark
255 Q24 instrument was used to run Pryosequencing assays according to the
256 manufacturer's instructions (Qiagen). Analyses of Pyrograms were conducted on the

257 PyroMark Q24 software (v 2.0.6., build 20; Qiagen). Seven CpGs representing 5
258 genes were chosen to provide a range of β-values. These demonstrated a strong
259 positive correlation between β-values and percentage methylation by bisulphite
260 sequencing (Spearman's r = 0.92, **Figure S2**).
261
262 Data analysis
263 Each array passed quality control assessment based on the performance of internal
264 array controls. Initial processing, probe type correction and assessment of array data
265 was conducted using the *minfi* package and SWAN.[23, 24] Probes with known SNPs
266 were removed. All CpGs for which one or more of the 22 samples displayed detection
267 *p* values > 0.05 (indicating an unreliable site) or presented with missing β-values were
268 excluded. The genomic inflation factor (λ, the ratio of the median of the observed
269 distribution of the test statistic to the expected median) was calculated using the
270 estlambda function of GenABEL.[25]
271
272 We filtered the data using criteria shown in **Figure 1** to identify differentially
273 methylated sites between GDM and healthy pregnancies. In step the first analysis, we
274 elected to use a minimum β-value difference of 0.05, in part to permit comparisons
275 with a recent report describing DNA methylation in placenta and umbilical cord blood

from GDM pregnancies also using the 450k array platform (step 1, Figure 1).[22] The genes identified as differentially methylated were obtained from the supplementary data of this particular publication. We also compared our data with a separate cohort of placenta and umbilical cord blood samples from GDM pregnancies.[17] We obtained their list of differentially methylated genes through personal communication with the corresponding author of the publication. Further filtering steps were applied to facilitate a more stringent analysis. To reduce the number of non-variable sites to improve the statistical power of subsequent analyses, we removed all sites with β-values ≥0.8 and ≤0.2 in all 22 samples (step 2, Figure 1). This is an approach that has been used by our group and as well as others.[41-44] As described previously by our group, we consider it a more robust methodology to remove from the data set CpGs that failed in any one of the samples, instead of eliminating specific failed CpGs from specific samples.[42] We retained only those CpGs which had a mean β-value difference of ≥0.2 (step 3, Figure 1). Finally we examined the absolute β-values in each matched pairs. We used a cut-off of ≥0.2 mean β-values difference to identify CpGs with considerable methylation differences.

Hierarchical clustering was performed utilising Genesis software (v1.7.6) using Euclidian distance and average linkage criteria.[45] Enrichment of gene ontology terms

and biological pathways within the genes associated with differentially methylated CpGs were assessed using DAVID online software. [26, 32]

Acknowledgement

The authors are grateful to Khaled Ismail, Angela Rooney and Emma Hubball who obtained the EFFECT-M samples. This work was supported by funding provided by the University Hospital of North Midlands Charity.

References

1. American Diabetes Association. Diagnosis and classification of diabetes mellitus. Diabetes Care 2011; 34:S62-S9.

2. National Collaborating Centre for Women's and Children's health. Diabetes in pregnancy. Management of diabetes and its complications from preconception to the postnatal period. NICE 2015; 3.

3. Gorgal R, Gonçalves E, Barros M, Namora G, Magalhães Â, Rodrigues T, et al. Gestational diabetes mellitus: A risk factor for non-elective cesarean section. Journal of Obstetrics and Gynaecology Research 2012; 38:154-9.

4. Schmidt MI, Duncan BB, Reichelt AJ, Branchtein L. Gestational diabetes mellitus diagnosed with a 2-h 75-g oral glucose tolerance test and adverse pregnancy

314 outcomes. Diabetes Care 2001; 24:1151-5.

315 5. Bellamy L, Casas JP, Hingorani AD, Williams D. Type 2 diabetes mellitus after
316 gestational diabetes: a systematic review and meta-analysis. The Lancet 2009;
317 373:1773-9.

318 6. Akinci B, Celtik A, Genc S, Yener S, Demir T, Secil M, et al. Evaluation of
319 postpartum carbohydrate intolerance and cardiovascular risk factors in women with
320 gestational diabetes. Gynecological Endocrinology 2011; 27:361-7.

321 7. Rivero K, Portal VL, Vieira M, Behle I. Prevalence of the impaired glucose
322 metabolism and its association with risk factors for coronary artery disease in women
323 with gestational diabetes. Diabetes Research and Clinical Practice 2008; 79:433-7.

324 8. Ling C, Del Guerra S, Lupi R, Rönn T, Granhall C, Luthman H, et al. Epigenetic
325 regulation of PPARGC1A in human type 2 diabetic islets and effect on insulin
326 secretion. Diabetologia 2008; 51:615-22.

327 9. Kulkarni SS, Salehzadeh F, Fritz T, Zierath JR, Krook A, Osler ME. Mitochondrial
328 regulators of fatty acid metabolism reflect metabolic dysfunction in type 2 diabetes
329 mellitus. Metabolism 2012; 61:175-85.

330 10. Yang BT, Dayeh TA, Volkov PA, Kirkpatrick CL, Malmgren S, Jing X, et al.
331 Increased DNA methylation and decreased expression of PDX-1 in pancreatic islets
332 from patients with type 2 diabetes. Molecular Endocrinology 2012; 26:1203-12.

11. Yang BT, Dayeh TA, Kirkpatrick CL, Taneera J, Kumar R, Groop L, et al. Insulin promoter DNA methylation correlates negatively with insulin gene expression and positively with HbA1c levels in human pancreatic islets. Diabetologia 2011; 54:360-7.

12. Hall E, Dayeh T, Kirkpatrick CL, Wollheim CB, Dekker Nitert M, Ling C. DNA methylation of the glucagon-like peptide 1 receptor (GLP1R) in human pancreatic islets. BMC medical genetics 2013; 14:76.

13. Ribel-Madsen R, Fraga MF, Jacobsen S, Bork-Jensen J, Lara E, Calvanese V, et al. Genome-wide analysis of DNA methylation differences in muscle and fat from monozygotic twins discordant for type 2 diabetes. PloS one 2012; 7:e51302.

14. Gluckman PD, Hanson MA, Buklijas T, Low FM. Epigenetic mechanisms that underpin metabolic and cardiovascular diseases. Nature Reviews Endocrinology 2009; 5:401-8.

15. Bouchard L, Thibault S, Guay SP, Santure M, Monpetit A, St-Pierre J, et al. Leptin gene epigenetic adaptation to impaired glucose metabolism during pregnancy. Diabetes Care 2010; 33:2436-41.

16. Bouchard L, Hivert MF, Guay SP, St-Pierre J, Perron P, Brisson D. Placental adiponectin gene DNA methylation levels are associated with mothers' blood glucose concentration. Diabetes 2012; 61:1272-80.

17. Ruchat SM, Houde AA, Voisin G, St-Pierre J, Perron P, Baillargeon JP, et al.

352 Gestational diabetes mellitus epigenetically affects genes predominantly involved in
353 metabolic diseases. Epigenetics 2013; 8:935-43.
354 18. Houde AA, Guay SP, Desgagné V, Hivert MF, Baillargeon JP, St-Pierre J, et al.
355 Adaptations of placental and cord blood ABCA1 DNA methylation profile to maternal
356 metabolic status. Epigenetics 2013; 8:1289-302.
357 19. El Hajj N, Pliushch G, Schneider E, Dittrich M, Müller T, Korenkov M, et al.
358 Metabolic programming of MEST DNA methylation by intrauterine exposure to
359 gestational diabetes mellitus. Diabetes 2013; 62:1320-8.
360 20. Lehnen H, Zechner U, Haaf T. Epigenetics of gestational diabetes mellitus and
361 offspring health: the time for action is in early stages of life. Molecular Human
362 Reproduction 2013; 19:415-22.
363 21. Ma RC, Tutino GE, Lillycrop KA, Hanson MA, Tam WH. Maternal diabetes,
364 gestational diabetes and the role of epigenetics in their long term effects on offspring.
365 Progress in Biophysics and Molecular Biology 2015; 118:55-68.
366 22. Finer S, Mathews C, Lowe R, Smart M, Hillman S, Foo L, et al. Maternal
367 gestational diabetes is associated with genome-wide DNA methylation variation in
368 placenta and cord blood of exposed offspring. Human Molecular Genetics 2015;
369 24:3021-9.
370 23. Aryee MJ, Jaffe AE, Corrada-Bravo H, Ladd-Acosta C, Feinberg AP, Hansen KD,

et al. Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. Bioinformatics 2014; 30:1363-9.

24. Maksimovic J, Gordon L, Oshlack A. SWAN: Subset-quantile within array normalisation for Illumina Infinium HumanMethylation450 BeadChips. Genome Biology 2012; 13:R44.

25. Aulchenko YS, Ripke S, Isaacs A, van Duijn CM. GenABEL: an R library for genome-wide association analysis. Bioinformatics 2007; 23:1294-6.

26. Huang DW, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nature Protocols 2008; 4:44-57.

27. Maglott D, Ostell J, Pruitt KD, Tatusova T. Entrez Gene: gene-centered information at NCBI. Nucleic Acids Res 2005; 33:D54-8.

28. Baron Gaillard CL, Pallesi-Pocachard E, Massey-Harroche D, Richard F, Arsanto JP, Chauvin JP, et al. Hook2 is involved in the morphogenesis of the primary cilium. Molecular Biology of the Cell 2011; 22:4549-62.

29. Krämer H, Phistry M. Genetic analysis of hook, a gene required for endocytic trafficking in drosophila. Genetics 1999; 151:675-84.

30. Haeseleer F, Jang GF, Imanishi Y, Driessen CAGG, Matsumura M, Nelson PS, et al. Dual-substrate specificity short chain retinol dehydrogenases from the vertebrate

390 retina. Journal of Biological Chemistry 2002; 277:45537-46.

391 31. Kanehisa M, Susumu G. KEGG: Kyoto encyclopedia of genes and genomes.

392 Nucleic Acids Research 2000; 28:27-30.

393 32. Huang DW, Sherman BT, Lempicki RA. Bioinformatics enrichment tools: paths

394 toward the comprehensive functional analysis of large gene lists. Nucleic Acids

395 Research 2009; 37:1-13.

396 33. Liu C, Guo LQ, Menon S, Jin D, Pick E, Wang X, et al. COP9 signalosome

397 subunit Csn8 is involved in maintaining proper duration of the G1 phase. Journal of

398 Biological Chemistry 2013; 288:20443-52.

399 34. Rolland T, Taşan M, Charloteaux B, Pevzner Samuel J, Zhong Q, Sahni N, et al.

400 A proteome-scale map of the human interactome network. Cell; 159:1212-26.

401 35. Johnson C, Marriott SJ, Levy LS. Overexpression of p101 activates PI3Kγ

402 signaling in T cells and contributes to cell survival. Oncogene 2007; 26:7049-57.

403 36. Shymanets A, Prajwal, Vadas O, Czupalla C, LoPiccolo J, Brenowitz M, et al.

404 Different inhibition of Gβγ-stimulated class IB phosphoinositide 3-kinase (PI3K)

405 variants by a monoclonal antibody. Specific function of p101 as a Gβγ-dependent

406 regulator of PI3Kγ enzymatic activity. Biochemical Journal 2015; 469:59-69.

407 37. Brock C, Schaefer M, Reusch HP, Czupalla C, Michalke M, Spicher K, et al.

408 Roles of Gβγ in membrane recruitment and activation of p110γ/p101 phosphoinositide

409 3-kinase γ. The Journal of Cell Biology 2003; 160:89-99.

410 38. Kohler C, Eriksson LG, Flood PR, Hardie JA, Okuno E, Schwarcz R. Quinolinic
411 acid metabolism in the rat brain. Immunohistochemical identification of
412 3-hydroxyanthranilic acid oxygenase and quinolinic acid phosphoribosyltransferase in
413 the hippocampal region. The Journal of Neuroscience 1988; 8:975-87.

414 39. Telkoparan P, Erkek S, Yaman E, Alotaibi H, Bayik D, Tazebay UH. Coiled-coil
415 domain containing protein 124 is a novel centrosome and midbody protein that
416 interacts with the Ras-guanine nucleotide exchange factor 1B and is involved in
417 cytokinesis. PloS one 2013; 8:e69289.

418 40. Fryer AA, Nafee TM, Ismail KM, Carroll WD, Emes RD, Farrell WE. LINE-1 DNA
419 methylation is inversely correlated with cord plasma homocysteine in man: a
420 preliminary study. Epigenetics 2009; 4:394-8.

421 41. Fryer AA, Emes RD, Ismail KM, Haworth KE, Mein C, Carroll WD, et al.
422 Quantitative, high-resolution epigenetic profiling of CpG loci identifies associations
423 with cord blood plasma homocysteine and birth weight in humans. Epigenetics 2011;
424 6:86-94.

425 42. Glossop JR, Nixon NB, Emes RD, Haworth KE, Packham JC, Dawes PT, et al.
426 Epigenome-wide profiling identifies significant differences in DNA methylation
427 between matched-pairs of T- and B-lymphocytes from healthy individuals. Epigenetics

428 2013; 8:1188-97.

429 43. Byun HM, Siegmund KD, Pan F, Weisenberger DJ, Kanel G, Laird PW, et al.
430 Epigenetic profiling of somatic tissues from human autopsy specimens identifies
431 tissue- and individual-specific DNA methylation patterns. Human Molecular Genetics
432 2009; 18:4808-17.

433 44. Glossop JR, Emes RD, Nixon NB, Packham JC, Fryer AA, Mattey DL, et al.
434 Genome-wide profiling in treatment-naive early rheumatoid arthritis reveals DNA
435 methylome changes in T- and B-lymphocytes. Epigenomics 2015 Nov 10. [Epub
436 ahead of print].

437 45. Sturn A, Quackenbush J, Trajanoski Z. Genesis: cluster analysis of microarray
438 data. Bioinformatics 2002; 18:207-8.

Table 1. Annotation for the 5 genes differentially methylated in 8 of 11 matched pairs, as determined by genome-wide DNA methylation analysis. *The official gene symbol, gene name and stated function were retrieved from the NCBI Gene database (accessed September 2015). **The absolute β-value difference range is the minimum to the maximum value of the individual absolute β-value differences for each differentially methylated CpG.

Gene symbol*	**Absolute β-value difference range****	**Gene name***	**Functional summary**
COPS8	0.05-0.84	Constitutive photomorphogenic homolog subunit 8	Regulator of multiple signaling pathways
PIK3R5	0.02-0.82	Phosphoinositide-3-kinase, regulatory subunit 5	Cell growth, proliferation, differentiation, motility, survival, and intracellular trafficking
HAAO	0.02-0.77	3-hydroxyanthranilate 3,4-dioxygenase	Catalyses the synthesis of quinolinic acid (QUIN), which is an excitotoxin that may participate in the pathogenesis of neurologic and inflammatory disorders
CCDC124	0.01-0.79	Coiled-coil domain containing 124	Cell cycle, cell division
C5orf34	0.01-0.77	Chromosome 5 open reading frame 34	Unknown, but sequence is

			conserved in chimpanzee, Rhesus monkey, dog, cow, mouse, rat, chicken, and zebrafish

Supplementary Table S1. Full annotated list of 100 differentially methylated CpGs.

CpG	UCSC_Refgene_Name	UCSC_Refgene_Group	UCSC_CPG_Islands_Name	Relation_to_UCSC_CpG_Island
cg03206401	TUBA3E	TSS200	chr2:130955619-130956274	Island
cg16322792	ZNF697	3'UTR	chr1:120165302-120166626	Island
cg17155524	ZFYVE28	Body	chr4:2305514-2305793	Island
cg21358336			chr17:6558221-6558441	Island
cg24686902			chr17:6558221-6558441	Island
cg24863815	NFIC	Body	chr19:3398628-3398935	Island
cg00684178	NEU4	5'UTR;1stExon	chr2:242754323-242754629	Shelf
cg02877261			chr4:186064047-186064614	Shelf
cg06684911	ATP8B3	Body	chr19:1795922-1797001	Shelf
cg08436396	LYPD5	TSS1500;5'UTR	chr19:44302665-44303176	Shelf
cg09101062	C5orf34	Body	chr5:43483519-43484555	Shelf
cg11331837			chr17:35165323-35165983	Shelf
cg12515659	FAM134B	Body	chr5:16616509-16617428	Shelf
cg13033971			chr13:46287282-46288214	Shelf
cg16569309			chr19:36266234-36266622	Shelf
cg17830140	POLRMT	Body	chr19:615691-623505	Shelf
cg18391209	CAPN8	Body	chr1:223741965-223744525	Shelf
cg18678716			chr5:195087-195323	Shelf
cg20976286	OCA2	Body	chr15:28050250-28050789	Shelf

CpG	UCSC_Refgene_Name	UCSC_Refgene_Group	UCSC_CPG_Islands_Name	Relation_to_UCSC_CpG_Island
cg21211688	ADAMTSL2	Body	chr9:136399367-136400274	Shelf
cg21927991	ZFAT	Body	chr8:135490786-135491086	Shelf
cg22274196			chr13:95953337-95954211	Shelf
cg24976563	DCAF11	Body	chr14:24583268-24584243	Shelf
cg25174111	MUS81	Body	chr11:65624495-65628596	Shelf
cg26864826			chr11:33757476-33758122	Shelf
cg01225004			chr14:101923575-101925995	Shore
cg02823329	PIK3R5	Body	chr17:8791470-8792004	Shore
cg03292225	TNNT3	Body	chr11:1958934-1959247	Shore
cg04131969	MYADML	Body	chr2:33952422-33952684	Shore
cg05237503	FBXO3	TSS1500	chr11:33795392-33796319	Shore
cg05305893	FGF11;CHRNB1	3'UTR;TSS1500	chr17:7348274-7348830	Shore
cg05918715	SHISA2	Body	chr13:26624725-26626265	Shore
cg06012903	PTPRN2	Body	chr7:157980786-157981462	Shore
cg06223162	GPR88	TSS200	chr1:101004471-101005885	Shore
cg07878625	ZNF783	Body	chr7:148978762-148979390	Shore
cg08332163	AIM1L	TSS200	chr1:26672445-26672650	Shore
cg08693140			chr7:6654745-6655860	Shore
cg09084244	CDK2AP1	TSS1500	chr12:123755246-123756408	Shore
cg09086151	HLA-DRB1	Body	chr6:32551851-32552331	Shore
cg14114910	MORN5	Body	chr9:124921950-124922170	Shore

CpG	UCSC_Refgene_Name	UCSC_Refgene_Group	UCSC_CPG_Islands_Name	Relation_to_UCSC_CpG_Island
cg15441831	CLDN4	TSS200	chr7:73245434-73246045	Shore
cg16995742	COPS8	TSS1500	chr2:237994004-237994876	Shore
cg18624102	FBXO27	TSS1500	chr19:39522264-39523227	Shore
cg21838924	CLDN4	TSS200	chr7:73245434-73246045	Shore
cg22996768			chr19:33717512-33717930	Shore
cg01105403				
cg01153376	MIR662;MSLN	3'UTR;TSS1500		
cg01835922				
cg01872988	DKFZp686A1627	TSS1500		
cg01979298				
cg02113055				
cg02389264				
cg02909570				
cg03129555				
cg03706056	SETD4	TSS1500		
cg04028540				
cg04497820				
cg05138546	KRT36	TSS200		
cg05515244	CDH5	5'UTR		
cg05531409	CPNE4	5'UTR		
cg05809586	KRTAP27-1	1stExon		

CpG	UCSC_Refgene_Name	UCSC_Refgene_Group	UCSC_CPG_Islands_Name	Relation_to_UCSC_CpG_Island
cg06002687				
cg06281714				
cg06407043				
cg06979386				
cg07240846	CAMK1D	Body		
cg07576186	PDHB	3'UTR		
cg08084984	XYLT1	Body		
cg08669168	SCFD2	Body		
cg08963013	LRRTM4	Body		
cg09284209				
cg10058204	FLJ37201	Body		
cg10701801	OSBPL9	TSS200		
cg10858640	SDK1	Body		
cg11047442				
cg11786587				
cg11957130	ATXN7L1	Body;5'UTR		
cg12342501				
cg12469381	CHN2	TSS200;Body		
cg13160852	LOC399959	Body		
cg13469425	TEC	Body		
cg14007688	DBH	Body		

CpG	UCSC_Refgene_Name	UCSC_Refgene_Group	UCSC_CPG_Islands_Name	Relation_to_UCSC_CpG_Island
cg14044669	C6orf10	Body		
cg14060113	CCDC124	3'UTR		
cg17174466	SPATS2L	Body		
cg17283620	HAAO	Body		
cg17738613	GPC5	Body		
cg17839758	C21orf29;KRTAP12-3	Body;TSS1500		
cg18584561	GREB1	TSS1500;5'UTR		
cg19248407	CUX1	Body		
cg19393008	KRT82	Body		
cg22274273				
cg22304519				
cg22436195				
cg24136292	INSC	Body		
cg24470466	HLA-DQA1	Body		
cg24534774				
cg25550823				
cg25673075				
cg27079096	OR52B4	TSS200		

446

447

448 **Supplementary Table S2**. Pathway analysis using DAVID. The gene ontology pathways and functional category terms, which were

449 overrepresented by the 66 genes identified by differential methylation (difference in β-values >0.2) between GDM and control cases, are ordered

450 by their *p* value. FDR= false discovery rate.

Category	Term	*p* Value	Fold Enrichment	FDR
KEGG_PATHWAY	hsa04514:Cell adhesion molecules (CAMs)	0.006	9.63	6.05
KEGG_PATHWAY	hsa04940:Type I diabetes mellitus	0.007	22.70	6.38
SP_PIR_KEYWORDS	Keratin	0.011	8.62	11.93
UP_SEQ_FEATURE	Sequence variant	0.016	1.23	18.80
GOTERM_BP_FAT	GO:0009405~Pathogenesis	0.016	118.67	20.37
SP_PIR_KEYWORDS	Polymorphism	0.018	1.24	19.00
GOTERM_CC_FAT	GO:0005882~intermediate filament	0.022	6.50	22.28
GOTERM_CC_FAT	GO:0045111~intermediate filament cytoskeleton	0.023	6.36	23.41
GOTERM_MF_FAT	GO:0030280~structural constituent of epidermis	0.028	68.33	29.11
SP_PIR_KEYWORDS	Alternative splicing	0.039	1.33	37.35
KEGG_PATHWAY	hsa04670: Leukocyte transendothelial migration	0.046	8.08	37.75

451

Supplementary Table S3. (A) Clinical characteristics of GDM and healthy pregnancy samples. (B) Summary data of the study population. C, control. G, GDM. BMI, body mass index. IVF, *in vitro* fertilisation. SD, standard deviation.

(A)

Sample ID	Age	BMI	Parity	Ethnicity	Smoking Status	Medication	Folic acid
G1	20	35	0	White British	Ex-smoker	None	Yes
C1	39	33	0	White British	Ex-smoker	None	Yes
G2	32	20	0	White British	Non-smoker	IVF medication prior to pregnancy	Yes
C2	22	20	1	White British	Non-smoker	None	Yes
G3	22	36	1	White British	Non-smoker	None	Yes
C3	29	38	1	White British	Non-smoker	Orlistat prior to pregnancy	Yes
G4	33	37	1	White British	Smoker	None	Yes
C4	25	34	1	White British	Smoker	None	Yes
G5	26	30	1	White British	Non-smoker	Amoxicillin prior to pregnancy	Yes
C5	26	30	0	White British	Non-smoker	None	Yes
G6	35	25	1	Asian or Asian British-Indian	Non-smoker	Thyroxine	Yes
C6	33	28	0	Asian or Asian British-Pakistani	Non-smoker	None	Yes
G7	38	27	0	White British	Non-smoker	None	Yes

C7	33	27	1	White British	Non-smoker	None	Yes
G8	40	35	3	White British	Non-smoker	Thyroxine	Yes
C8	31	34	2	White British	Non-smoker	Medication for hay fever	Yes
G9	32	39	0	Western Europe	Non-smoker	None	Yes
C9	30	44	1	White British	Non-smoker	None	Yes
G10	35	27	1	Asian or Asian-British Indian	Non-smoker	Antibiotics	Yes
C10	33	31	4	Asian or Asian British-Indian	Non-smoker	Antibiotics	Yes
G11	36	24	1	White British	Non-smoker	Medication for vertigo	Yes
C11	35	26	2	White British	Non-smoker	None	Yes

(B)

Clinical Characteristics	**GDM (*n* = 11)**	**Healthy (*n* = 11)**
Age (year, mean ± SD)	31.7 ± 6.4	30.6 ± 4.8
BMI (median, interquartile range)	30 (25-36)	31 (27-34)
Parity (median, interquartile range)	1 (0-1)	1 (0-2)

458 **Legends**

459 **Figure 1.** Filtering criteria for the identification of CpGs differentially methylated
460 between GDM and normal pregnancies. The starting number of CpGs (484,273) was
461 derived through the removal of CpGs with high detection values (p >0.05) and those
462 with missing β-values in any one of the 22 samples, as described in the Materials and
463 Methods. Horizontal line denotes additional filtering steps. *According to Finer *et al*.
464 criteria.[22] **According to Ruchat *et al*. criteria.[17] GDM, gestational diabetes. SD,
465 standard deviation.

466

467 **Figure 2.** Venn diagrams illustrating comparison of genes differentially methylated in
468 GDM using maternal blood with those identified in cord blood and placenta of GDM
469 affected pregnancies from the cohorts of (A) Finer *et al*.[22] and (B) Ruchat *et al*.[17],
470 respectively. The genes from our dataset that were common with the other study are
471 shown in dark gray shading. Genes identified as differentially methylated in Finer *et*
472 *al*.[22] were obtained from Supplementary file 2 of the published article, while the list of
473 differentially methylated genes identified by Ruchat *et al*.[17] was kindly provided
474 through personal communication with the corresponding author of Ruchat *et al*.[17]

475

476 **Figure 3**. Heatmap and dendrograms showing clustering[45] for the 100 CpGs
477 identified as differentially methylated (mean difference in β-values >0.2) between
478 GDM and healthy pregnancies. DNA methylation across the 100 sites in each of the
479 samples was analysed by hierarchical clustering using the Euclidean distance and
480 average linkage criteria. Each row represents an individual CpG site and each
481 column a different sample. Healthy controls and GDM samples are shown by the
482 green and red bars, respectively. Slide type is also shown with slide 1 in green and
483 slide 2 in red. Color gradation from yellow to blue represents low to high DNA
484 methylation respectively, with β-values ranging from 0 (no methylation; yellow) to 1
485 (complete methylation; blue). GDM, gestational diabetes.
486
487 **Supplementary Figure S1**. Characteristics of the 100 CpGs identified as differentially
488 methylated using genome-wide 450k arrays. (A) Frequency of CpG sites according to
489 their genomic location. (B) DNA methylation in GDM versus control samples plotted
490 by genomic location. The Illumina HumanMethylation450 BeadChip-Array annotations
491 are used as the basis for gene regions. Data are presented as mean ± standard
492 deviation. (C) Frequency of CpG sites according to their relationship with CpG islands.
493 (D) DNA methylation in GDM versus control samples plotted by relationship with CpG
494 islands. TSS proximal promoter defined as 200 bp (TSS200) or 1,500 bp (TSS1500)

upstream of the transcription start site. UTR, untranslated region. GDM, gestational diabetes.

Supplementary Figure S2. Technical validation of 450k BeadChip-Array data by sodium bisulfite Pyrosequencing. Correlation of DNA methylation as measured by array β-value and by bisulfite Pyrosequencing for 166 individual sites in the participants. Spearman's r = 0.92. (7 CpGs from 5 separate genes were selected to cover across the range of β-values: *AHRR*: cg23576855, *IGF2*: cg27331871, *Mir886*: cg26896946, cg26328633, *PM20D1*: cg07167872, cg24503407, *TCF7L2*: cg00159523).